



**FUNDAÇÃO UNIVERSIDADE FEDERAL DE RONDÔNIA –**  
**UNIR *CAMPUS* JI-PARANÁ**  
**Departamento de Matemática e Estatística**



**JOSÉ EDUARDO AGULHARE**

**ANÁLISE MULTIVARIADA DOS ACIDENTES DE TRÂNSITO NA BR-364**  
**NO MUNÍCIPIO DE JI-PARANÁ - RO**

**JI-PARANÁ – RO**  
**DEZEMBRO – 2017**

JOSÉ EDUARDO AGULHARE

**ANÁLISE MULTIVARIADA DOS ACIDENTES DE TRÂNSITO DA BR-364  
NO MUNÍCIPIO DE JI-PARANÁ - RO**

Trabalho de Conclusão de Curso  
apresentado ao Departamento de  
Matemática e Estatística, da Fundação  
Universidade Federal de Rondônia (UNIR)  
Campus de Ji-Paraná como parte dos  
requisitos para obtenção do título de  
Bacharel em Estatística.

Orientadora: Prof.<sup>a</sup> Dr<sup>a</sup>. Roziane Sobreira  
dos Santo

**JI-PARANÁ – RO  
DEZEMBRO – 2017**

Dados Internacionais de Catalogação na Publicação  
Fundação Universidade Federal de Rondônia  
Gerada automaticamente mediante informações fornecidas pelo(a) autor(a)

---

A284a Agulhare, José Eduardo.

Análise Multivariada dos acidentes de trânsito na BR-364 no município de Ji-Paraná - RO / José Eduardo Agulhare. -- Ji-Paraná, RO, 2017.

43 f.

Orientador(a): Prof.<sup>o</sup> Dra. Roziane Sobreira dos Santos

Trabalho de Conclusão de Curso (Graduação em Estatística) - Fundação Universidade Federal de Rondônia

1.Acidentes de Trânsito. 2.Polícia Rodoviária Federal. 3.Ji-Paraná.  
4.Análise Fatorial. I. Santos, Roziane Sobreira dos. II. Título.

CDU 519.233.4

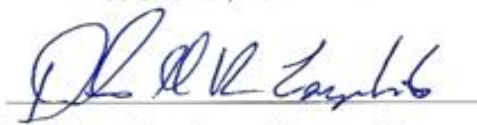
## ATA DA SESSÃO DE APRESENTAÇÃO E DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

Aos 13 (três) dias do mês de dezembro de 2017, realizou-se na Sala Laboratório de Estatística 1, no Campus de Ji-Paraná, a Sessão de Apresentação e Defesa do Trabalho de Conclusão de Curso (TCC), com o título “**Análise Multivariada dos Acidentes de Trânsito na BR-364 no Município de Ji-Paraná**”, apresentado pelo acadêmico **José Eduardo Agulhare**. Os trabalhos foram instalados às 15h34min pela presidente da Banca Examinadora, aprovada pelo Departamento e constituída por: Prof<sup>o</sup>. **Gd. Bruno Soares de Castro** (Universidade Federal de Rondônia - Departamento de Matemática e Estatística), Prof<sup>o</sup>. **Dr. Dilson Henrique Ramos Evangelista** (Universidade Federal de Rondônia - Departamento de Matemática e Estatística) e Prof<sup>a</sup> **Dr<sup>a</sup>. Roziane Sobreira dos Santos** (Universidade Federal de Rondônia - Departamento de Matemática e Estatística), sendo esta a orientadora do acadêmico. A Banca Examinadora, tendo decidido aceitar o Trabalho de Conclusão de Curso, após a apresentação passou à arguição pública do acadêmico. Encerrando os trabalhos de arguição às 16h51min. A Banca Examinadora deu parecer final **Aprovado**, com a nota 90 (**noventa**), resultado da média aritmética das notas individuais atribuídas pelos membros da Banca Examinadora. Proclamado o resultado final pela presidente da Banca, foram encerrados os trabalhos. Para constar, lavrou-se a presente Ata, que é assinada pelos membros da Banca Examinadora e pelo acadêmico.

Ji-Paraná, 13 de dezembro de 2017.



**Gr. Bruno Soares de Castro**  
Prof<sup>o</sup>. Membro da Banca



**Dr<sup>o</sup>. Dilson Henrique Ramos Evangelista**  
Prof<sup>o</sup>. Membro da Banca



**Dr<sup>a</sup>. Roziane Sobreira dos Santos**  
Prof<sup>a</sup> Orientadora e Pres. Banca



**José Eduardo Agulhare**  
Acadêmico

*Dedico este trabalho a Deus, também a toda  
minha família, aos meus amigos que  
acreditam em mim e a todos que me  
apoiam durante minha trajetória  
acadêmica. Obrigado!*

## **AGRADECIMENTOS**

Reunir aqui todos os que me ajudaram e me apoiaram durante todo o curso e também antes dele, creio que seja a parte mais difícil deste trabalho. Entretanto, agradeço muito a Deus em primeiro lugar por me conceder vida e saúde por todo este tempo.

Tenho também eterno agradecimento aos meus pais, Edelsina e José, e meu irmão Benito que tiveram contribuição direta e indiretamente em minha formação. Acredito que se não fosse por eles nada disto seria possível. Toda a minha família sempre agindo com presteza em me ajudar no que fosse preciso. Meu irmão, Benito, que sempre me ajudou da forma que podia mesmo sem saber o fim de tudo isso e a pessoa que eu tanto amo Edilaine Alves. Obrigado por tudo!

Não posso deixar de prestar meus agradecimentos aos meus grandes profissionais e docentes da UNIR Campus de Ji-paraná, que não mediram esforços em transmitir seus conhecimentos ao longo do curso, obrigado a todos sem exceção. Agradeço a minha estimada orientadora professora Dr.<sup>a</sup> Roziane Sobreira e meu professor Bruno Soares pelos excelentes ensinamentos e lições. Também agradeço aos professores Dr. Dilson, Dr. Nério, ao Me. Josivam que foram pessoas que transmitiram seus conhecimentos a mim. Sou muito grato por tudo que fizeram.

Agora quero agradecer aos meus amigos de faculdade, em especial o Emanuel, Luis, Tuanny, Weverson e a Maína que trago desde o início do curso, resistindo aos desafios, cada um com suas dificuldades assim como eu e a toda minha turma. Obrigado por estarem junto comigo durante este tempo. A todos os outros amigos que também gostaria de citar, mas o espaço é curto, resumo agradecendo aos meus amigos de Ji-paraná, em especial aos meus amigos de trabalho Lucas Henrique, Pablo, Thiago, Alex, Wellison e Kaio e também ao meu Patrão Wesley aos meus amigos de trabalho foi muito bom poder contar com todos vocês.

Muito obrigado a todos e todas, mesmo os que não foram citados, todos vocês foram importantes para mim.

"Torture seus dados até que eles confessem  
o que você deseja comprovar"

Rosângela Leal

## RESUMO

AGULHARE, J. E. **Análise Multivariada dos Acidentes de Trânsito da BR-364 no Município de Ji-Paraná – RO.** Monografia (Bacharelado em Estatística) Departamento de Matemática e Estatística, Universidade Federal de Rondônia, 2017.

Os acidentes de trânsito se tornaram uma das principais causas de mortes na atualidade, fazendo com que esta se caracterize uma das principais preocupações no habitual da vida urbana. Este trabalho faz um estudo dos acidentes ocorridos na BR-364 no período de janeiro de 2015 a dezembro de 2016 no município de Ji-paraná RO, buscando caracterizar os diversos tipos de acidentes com as condições de pista, condições meteorológicas, quilômetro de ocorrência, entre outras. Foi utilizada a técnica de Análise Fatorial para encontrar os possíveis fatores dentro dos registros de acidentes, utilizando dentro da técnica um recurso de rotação dos fatores para a obtenção da interpretabilidade de cada fator. Os registros utilizados neste trabalho estão dispostos em planilhas, no site da Polícia Rodoviária Federal. Para proceder com a análise fatorial dos acidentes de trânsito na BR-364, foi necessário organizar e separar os registros ocorridos no município de Ji-paraná RO, quantificando e dando uma ordem numérica as observações de cada variável, devido ao fato destas planilhas serem compostas de dados qualitativos e quantitativos. Ao todo foram 416 acidentes ocorridos no período estudado. Todo esse processo se fez necessário para, explicar de forma resumida a variabilidade do banco de dados através das cargas fatoriais de cada fator. A Análise Fatorial foi realizada duas vezes sendo que a primeira aplicação resultou na eliminação de algumas variáveis que pouco explicavam os fatores obtidos, e apenas a segunda Análise Fatorial foi interpretada. Com os resultados preliminares pôde-se verificar os trechos de maior ocorrência de acidentes localizados dentro do perímetro urbano de Ji-paraná com destaque para o trecho da ponte do Rio Machado que liga o primeiro distrito ao segundo distrito do município. Após a rotação dos fatores foram identificados três fatores definimos como: fator diário meteorológico, fator decorrente do dia da semana e fator estrutural do quilômetro de ocorrência.

**Palavras-Chave:** Acidentes de Trânsito, Polícia Rodoviária Federal, Ji-paraná, Análise Fatorial.



## **ABSTRACT**

AGULHARE, J. E. **Multivariate Analysis of BR-364 Traffic Accidents in the Municipality of Ji-Paraná - RO.** Monography (Bachelor of Statistics) Department of Mathematics and Statistics, Federal University of Rondônia, 2017.

Traffic accidents have become one of the main causes of death today, making this one of the main concerns in the usual urban life. This study makes a study of the accidents occurred in the BR-364 from January 2015 to December 2016 in the municipality of Ji-paraná RO, seeking to characterize the different types of accidents with the piste conditions, meteorological conditions, kilometer of occurrence, among others. The factorial analysis technique was used to find the possible factors within the accident records, using within the technique a resource of rotation of the factors to obtain the interpretability of each factor. In order to proceed with the factorial analysis of traffic accidents in BR-364, it was necessary to organize and separate the records that occurred in the municipality of Ji-paraná RO, discretizing and giving numerical order the observations of each variable, due to the fact that these worksheets are composed of qualitative and quantitative data. In all, 416 accidents occurred during the study period. All this process was necessary to explain in a summarized way the variability of the database through the factorial loads of each factor. The Factorial Analysis was performed twice and the first application resulted in the elimination of some variables that did not explain the factors obtained, and only the second Factor Analysis was interpreted. With the preliminary results, it was possible to verify the most frequent occurrences of accidents located within the urban perimeter of Ji-paraná, with emphasis on the stretch of the Machado River bridge that links the first district to the second district of the municipality. After the rotation of the factors were identified three factors defined as: daily meteorological factor, factor arising from the day of the week and structural factor of the kilometer of occurrence.

**Keywords:** Traffic Accidents, Federal Highway Police, Ji-paraná, Factor Analysis.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1.</b> Gráfico de frequência de acidentes entre os quilômetros de 310 ao 370.....	30
<b>Figura 2.</b> Teste Scree para análise de componentes .....	34
<b>Figura 3.</b> Gráfico da comunalidade e variância especificada.....	36
<b>Figura 4.</b> Projeção dos escores nos planos fatoriais com relação a variável uso do solo .....	39
<b>Figura 5.</b> Projeção dos escores nos planos fatoriais com relação a variável sentido da via.....	40

## LISTA DE TABELAS

<b>Tabela 1.</b> Tabela de ordenação das variáveis .....	29
<b>Tabela 2.</b> Descrição das variáveis .....	32
<b>Tabela 3.</b> Medida de adequação de cada variável na Primeira aplicação da Análise Fatorial .....	33
<b>Tabela 4.</b> Medida de adequação de cada variável na segunda aplicação da Análise Fatorial .....	34
<b>Tabela 5.</b> Rotação ortogonal da matriz fatorial. ....	35
<b>Tabela 6.</b> Interpretabilidade dos fatores a partir do critério de rotação Varimax.....	37

## LISTA DE EQUAÇÕES

<b>Equação 1.</b> Modelo Fatorial.....	18
<b>Equação 2.</b> Modelo fatorial em notação matricial .....	18
<b>Equação 3.</b> Suposição de que todos os fatores são correlacionados .....	19
<b>Equação 4.</b> Suposição de que todos os erros têm médias iguais à zero .....	20
<b>Equação 5.</b> Modelo ortogonal assumido pela matriz de correlação teórica reparametrizada .....	20
<b>Equação 6.</b> Variância total a proporção explicada pelo fator F.....	21
<b>Equação 7.</b> Estatística do teste de esfericidade de Bartlett.....	22
<b>Equação 8.</b> Índice de KMO.....	22
<b>Equação 9.</b> Índice de MSA.....	23
<b>Equação 10.</b> Matriz de parâmetros constantes .....	24
<b>Equação 11.</b> Matriz dos erros.....	24
<b>Equação 12.</b> Matriz de correlação amostral .....	25
<b>Equação 13.</b> Aproximação para a matriz de parâmetros constantes .....	25
<b>Equação 14.</b> Construção da matriz dos erros.....	25
<b>Equação 15.</b> Aproximação da matriz de correlação amostral original .....	25
<b>Equação 16.</b> Matriz Residual.....	25
<b>Equação 17.</b> Rotação dos fatores pelo critério Varimax.....	26
<b>Equação 18.</b> Estimação dos escores pelo método dos mínimos quadrados ponderados .....	27

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>14</b>
<b>1.1 OBJETIVOS .....</b>	<b>15</b>
1.1.1 OBJETIVO GERAL .....	15
1.1.2 OBJETIVOS ESPECÍFICOS .....	15
<b>2. FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>16</b>
2.1 ACIDENTES DE TRÂNSITO .....	16
2.2 ANÁLISE FATORIAL.....	16
2.3 ANÁLISE FATORIAL EXPLORATÓRIA .....	17
2.3.1 MODELO DE FATORES ORTOGONAL .....	19
2.3.2 TESTE DE ESFERICIDADE DE BARTLETT .....	21
2.3.3 MEDIDAS DE ADEQUAÇÃO.....	22
2.3.4 ESTIMAÇÃO DO NÚMERO DE FATORES $M$ .....	23
2.3.5 MÉTODO DE ESTIMAÇÃO DAS MATRIZES $L_{pxm}$ $\Psi_{pxp}$ .....	24
2.3.6 ROTAÇÃO DOS FATORES .....	25
2.3.7 CRITÉRIO DE ROTAÇÃO VARIMAX.....	26
2.3.8 ESTIMAÇÃO DOS ESCORES (MÍNIMOS QUADRADOS PONDERADOS) ...	26
<b>3. METODOLOGIA.....</b>	<b>27</b>
4.1 APLICAÇÃO DA ANÁLISE FATORIAL EXPLORATÓRIA .....	31
4.1.2 DETERMINAÇÃO DO NÚMERO DE FATORES.....	34
4.1.3 INTERPRETAÇÃO DOS FATORES.....	36
4.1.4 VALIDAÇÃO DO MODELO .....	38
<b>5. CONSIDERAÇÕES .....</b>	<b>40</b>
<b>6. CONCLUSÃO.....</b>	<b>41</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>42</b>

## 1. INTRODUÇÃO

Os acidentes de trânsito se tornaram uma das principais causas de mortes na atualidade, fazendo com que esta se caracterize uma grande preocupação no habitual da vida urbana, tanto para as autoridades e órgãos responsáveis, quanto para a população em geral. Sendo assim tal preocupação tem ganhado destaque em países desenvolvidos e subdesenvolvidos, que buscam uma redução nos danos causados por esse tipo de fatalidade, tanto danos materiais, quanto os danos pessoais (BERGAMASCHI, 2010).

Estudos divulgados pela Organização Mundial da Saúde (OMS, 2015) revelam que acidentes de trânsito no mundo, vitimam 1,25 milhão de pessoas por ano, acumulando óbitos principalmente em países pobres. Segundo a OMS, os países de baixa ou média renda acumulam 90% das mortes no trânsito enquanto somam 54% dos veículos do mundo.

O Brasil apresenta uma taxa de 23,4 mortes no trânsito para cada 100 mil habitantes, de acordo com estimativas da OMS, divulgadas na 68ª Assembleia Mundial da Saúde, em Genebra na Suíça (OMS, 2015). O país apresenta o quarto pior desempenho entre os países de América Latina, atrás de Belize, República Dominicana e Venezuela – campeã de acidentes na região com 45,1 mortes por 100 mil habitantes.

Um relatório divulgado pelo Instituto de Pesquisa Econômica Aplicada (IPEA, 2015) afirma que os acidentes de trânsito no Brasil tendem a se agravar, devido à expansão da frota de veículos automotores que o país vem vivendo desde o final do século passado. Implicando no aumento do fluxo de veículos nos grandes centros e rodovias, o que reflete diretamente na deterioração dessas vias de circulação. Além de pressionar os índices de acidentes nas rodovias do país, em especial as rodovias federais fiscalizadas pela Polícia Rodoviária Federal (PRF).

Sendo assim, os resultados obtidos neste trabalho ficarão à disposição do órgão responsável pela via, afim de que a partir deles possa-se analisar e identificar variáveis que mais contribuam para a carga fatorial na caracterização dos acidentes de trânsito ocorridos na BR-364 no município de Ji-Paraná/RO, identificando os principais pontos

de ocorrência de acidente e possíveis causas dos mesmos, para que se busquem medidas a fim de minimizar os diversos problemas, causados pelos acidentes na localidade em estudo.

Analisar informações de fenômenos que envolvam uma grande quantidade de variáveis relacionadas ao processo em estudo, exige a aplicação das técnicas da estatística multivariada. A utilização da Análise Fatorial neste trabalho busca agrupar e caracterizar as variáveis correlacionadas em fatores, resumindo assim a grande quantidade de variáveis envolvidas nos acidentes, além de identificar os pontos da rodovia com as principais causas e tipos de acidentes (MINGOTI, 2013).

Nesse sentido a Análise Fatorial se mostra a técnica multivariada mais adequada para buscar relações entre as variáveis registradas no Boletim de Acidentes de Trânsito (BAT) pela PRF na BR-364 em Ji-paraná. Os resultados obtidos através da análise fatorial nessas variáveis que poderão determinar quais os fatores latentes que mais contribuem para as causas desses acidentes, e assim, o órgão responsável pela via poderá desenvolver e aplicar medidas de correção em trechos específicos da rodovia.

## **1.1. OBJETIVOS**

### **1.1.1 OBJETIVO GERAL**

Aplicar análise estatística multivariada para descrever e caracterizar os diversos tipos de acidentes com as condições de pista, condições meteorológicas, quilômetro de ocorrência, entre outras na BR-364 no trecho que corta o município de Ji-paraná.

### **1.1.2 OBJETIVOS ESPECÍFICOS**

- a. Explicar de forma resumida a variabilidade do banco de dados.
- b. Empregar a técnica de análise fatorial para encontrar os fatores latentes dentro dos registros de acidentes.
- c. Utilizar critério de rotação dos fatores para a obtenção da interpretabilidade de cada fator.

- d. Encontrar e interpretar a carga fatorial dos fatores.
- e. Analisar os escores fatoriais para o tipo de localidade e sentido da via na ocorrência de acidentes.

## **2. FUNDAMENTAÇÃO TEÓRICA**

### **2.1 ACIDENTES DE TRÂNSITO**

Um acidente de trânsito pode ser definido como um acontecimento independente do desejo do homem, causado por uma força externa, que atua inesperadamente e deixa danos pessoais e/ou materiais, envolvendo um ou mais veículos em circulação, normalmente por uma via para trânsito de veículos, podendo ser o veículo motorizado ou não (GOLD, 1998).

O Conselho de Trânsito Brasileiro (CTB) define rodovia como uma via rural pavimentada. Entretanto, toma-se para esse trabalho a definição de rodovia como sendo uma via rural ou urbana, aberta à circulação pública, por onde transitam veículos, pessoas e animais, compreendendo a pista, o acostamento, e o canteiro central e que sejam fiscalizadas pela PRF (MINERVI, 2013).

Essa definição se enquadra melhor no contexto deste trabalho, pelo fato de a área estudada abranger boa parte do trecho da BR-364 em zona urbana. Tendo em vista que a definição de rodovia tomada pelo CTB, limita o real sentido de rodovia nesse trabalho. Além do fato de que no Brasil, as rodovias federais, em sua grande maioria atravessam grandes centros urbanos ao longo de sua extensão, assim, quando isso acontece, descaracteriza uma definição estrita de via rural.

### **2.2 ANÁLISE FATORIAL**

A análise fatorial foi desenvolvida inicialmente por Charles Spearman, que propôs a atual modelagem da estrutura fatorial, para estudar correlações entre escores de testes de estudantes em algumas áreas da educação (SPEARMAN, 1904). E que atualmente, se tornou uma das técnicas multivariadas mais conhecidas e utilizadas quando o pesquisador não tem um total conhecimento das relações de dependência entre as variáveis, tão pouco o que tais fatores representam (MINGOTI, 2013).



Segundo Johnson e Wichern (2002), o propósito da análise fatorial é descrever, um possível, relacionamento da covariância entre uma grande quantidade de variáveis e assim identificar variáveis não observáveis, chamadas de fatores. A modelagem da estrutura fatorial supõe um agrupamento de variáveis de acordo com suas correlações. Esse agrupamento ocorre quando, as variáveis do grupo particular são altamente correlacionadas entre si, em consequência tem menores correlações com variáveis dos demais grupos.

Diversos trabalhos podem ser citados pela aplicação de técnicas estatísticas multivariadas que buscavam encontrar a problemática dos acidentes de trânsito. Entre eles, Balbo et al. (2010) e Sehaber (2013) aplicaram análise fatorial exploratória buscando encontrar relações internas em de registros de acidentes de trânsito com o objetivo de agrupar variáveis relacionadas que pudessem caracterizar melhor as causas de ocorrência dos acidentes na rodovia. Minervi (2013) aplica a Análise Fatorial no conjunto de variáveis contidas no Boletim de Acidentes de Trânsito (BAT), para obter fatores existentes e reduzir o quantitativo de variáveis contidas no BAT.

A análise fatorial pode ser trabalhada de forma exploratória ou confirmatória. Para Hair et al. (2009) os pesquisadores usam a análise fatorial explanatória buscando encontrar os fatores extraídos das variáveis originais do banco de dados em estudo, considerando o que os dados tem a oferecer. Já na análise fatorial confirmatória, o pesquisador tem ideias de um modelo inicial hipotético baseado em suporte técnico ou em pesquisas anteriores, e se busca avaliar o grau em que os dados satisfazem a estrutura esperada confirmando se este modelo de relacionamento proposto está correto ou não. Neste trabalho iremos aplicar apenas a análise fatorial exploratória.

### 2.3 ANÁLISE FATORIAL EXPLORATÓRIA

Segundo Mingoti (2013), seja  $X_{pxl}$  um vetor aleatório, com n-variáveis observáveis, com um vetor de média  $\mu$ , onde,  $\mu = (\mu_1, \mu_2, \dots, \mu_n)$ , matriz de covariância  $\Sigma_{pxp}$  e matriz de correlação  $P_{pxp}$ . Sejam  $Z_i = [(X_1 - \mu_i)/\sigma_i]$  variáveis originais padronizadas, onde  $\mu_i$  e  $\sigma_i$  representem respectivamente a média e o desvio padrão da

variável  $X_i$ ,  $i = 1, 2, \dots, p$ . Neste caso, a matriz de correlação  $P$  é a matriz de covariâncias do vetor aleatório  $Z = (Z_1, Z_2, \dots, Z_p)'$ .

O modelo fatorial construído a partir da matriz de correlação teórica  $P_{p \times p}$  é formalizado conforme as equações dadas em 1 e 2 abaixo:

$$Z_1 = L_{11}F_1 + L_{12}F_2 + \dots + L_{1m}F_m + \varepsilon_1 \quad (1)$$

$$Z_2 = L_{21}F_1 + L_{22}F_2 + \dots + L_{2m}F_m + \varepsilon_2$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

$$Z_p = L_{p1}F_1 + L_{p2}F_2 + \dots + L_{pm}F_m + \varepsilon_p$$

Sendo  $m \leq p$ .

O modelo (1), em notação matricial é expresso por:

$$D(X - \mu) = LF + \varepsilon \quad (2)$$

Onde,

$$(X - \mu)_{p \times 1} = \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix}, \varepsilon_{p \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}, F_{m \times 1} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, L_{p \times m} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{bmatrix},$$

$$D_{p \times p} = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & \dots & 0 \\ 0 & 1/\sigma_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1/\sigma_p \end{bmatrix}$$

Assim,  $F_{m \times 1}$ , é um vetor aleatório com m fatores, também chamados de variáveis latentes, que descrevem os elementos da população estudada e não são observáveis,

$1 \leq m \leq p$ , ou seja não podem ser medidos a priori. Nesse sentido o modelo de análise fatorial assume que as variáveis  $Z$  estão correlacionadas linearmente com novas variáveis aleatórias  $F_j$ ,  $j = 1, 2, \dots, m$ , que deverão ser analisadas e identificadas pelo pesquisador. O Vetor  $\varepsilon_{pxl}$  é um vetor de erros aleatórios correspondentes aos erros de medidas e variação de  $Z_i$ , que não é explicada pelos fatores comuns  $F_j$ ,  $j = 1, 2, \dots, m$ , incluídos no modelo. O coeficiente  $l_{ij}$ , é o coeficiente da  $i$ -ésima variável padronizada  $Z_i$ , comumente chamado de peso ou Carga Fatorial (CF), no  $j$ -ésimo fator  $F_j$ , e representa o grau de relacionamento linear entre  $Z_i$  e  $F_j$ ,  $j = 1, 2, \dots, m$ . Os coeficientes  $l_{ij}$  ou cargas fatoriais, ficam armazenados na matriz  $L_{pxm}$ , que é uma matriz de parâmetros (constantes) a serem estimados. Assim de acordo com o modelo linear (1), as informações das  $n$ -variáveis originais padronizadas  $(Z_1, Z_2, \dots, Z_n)$  estarão sendo representadas por  $(p + m)$  variáveis aleatórias não observáveis, ou seja,  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$  e  $(F_1, F_2, \dots, F_m)$ .  $(p + m)$ .

### 2.3.1 MODELO DE FATORES ORTOGONAL

Algumas suposições são necessárias para que se possa operacionalizar a estimação do modelo em (1), Supõe-se que:

(i)  $E[F_{mxl}] = 0$ . Isso implica que  $E[F_j] = 0$ ,  $j = 1, 2, \dots, m$ , ou seja esta suposição diz que todos os fatores tem média igual a zero, consequentemente;

$$(ii) \quad Var[F_{mxm}] = I_{mxm} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \vdots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (3)$$

ou seja, todos os fatores  $F_j$  são não correlacionados e têm variâncias iguais a 1;

(iii)  $E[\varepsilon_{pxl}] = 0$ , o que implica que  $E[\varepsilon_j] = 0$ ,  $j = 1, 2, \dots, p$ , ou seja, todos os erros têm médias iguais a zero;

$$(iv) \quad Var[\varepsilon_{pxp}] = \Psi_{pxp} = \begin{bmatrix} \Psi_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Psi_2 & \cdots & \vdots \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \Psi_p \end{bmatrix} \quad (4)$$

ou seja,  $Var[\varepsilon_j] = \Psi_j$  e  $Cov(\varepsilon_i, \varepsilon_j) = 0$ ,  $\forall i \neq j$ , o que significa dizer que os erros são não correlacionados entre si e não necessariamente têm a mesma variância;

(v) Os vetores  $\varepsilon_{pxl}$  e  $F_{mxl}$  são independentes. Portanto,  $Cov(\varepsilon_{pxl}, F_{mxl}) = E(\varepsilon F') = 0$ .

A suposição (v) implica dizer que os vetores  $\varepsilon$  e  $F$  representam duas fontes de variação distintas, relacionadas às variáveis padronizadas  $Z_i$ , não havendo qualquer relacionamento entre estas fontes de informação. Um modelo fatorial com as suposições de (i) à (v) é chamado de ortogonal, onde a ortogonalidade refere-se ao fato de que os  $m$  fatores são ortogonais entre si. Uma consequência imediata das suposições acima está relacionada com a estrutura da matriz de correlação teórica  $P_{pxp}$ . Quando o modelo ortogonal é assumido, a matriz  $P_{pxp}$  pode ser reparametrizada na forma:

$$P_{pxp} = LL' + \Psi \quad (5)$$

isso vem do fato de que:

$$P_{pxp} = Var(Z) = Var(LF + \varepsilon) = Var(LF) + Var(\varepsilon) = LIL' + \Psi$$

onde  $I$  é a matriz identidade de dimensão  $pxp$ . O objetivo da análise fatorial é encontrar as matrizes  $L_{pxm}$  e  $\Psi_{pxp}$  que possam representar a matriz  $P_{pxp}$  para um dado valor de  $m$ , menor que o número de variáveis originais  $n$ . Infelizmente, existem muitas matrizes de correlação  $P_{pxp}$  que não podem ser decompostas na forma  $LL' + \Psi$  para um valor  $m$  muito menor que  $n$  (JOHNSON; WICHERN, 2002).

As implicações da decomposição (5) são:

$$(i) \quad Var(Z_i) = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2 + \Psi_i = h_i^2 + \Psi_i, \quad \text{onde} \quad h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2, \\ i = 1, 2, \dots, p.$$

isso significa que a variância de  $Z_i$  é decomposta em duas partes. A primeira, denotada por  $h_i^2$ , é a variabilidade de  $Z_i$  explicada pelos  $m$  fatores incluídos no modelo fatorial. Essa parte da variabilidade é chamada de “comunalidade”. A segunda, denotada por  $\Psi_i$ , é a parte da variabilidade de  $Z_i$  associada ao erro aleatório  $\varepsilon$ , o qual é específico de cada variável  $Z_i$ , parte essa chamada de “unidade” ou “variabilidade específica”. Como  $Z_i$  têm variâncias iguais a 1, segue que:

$$(i) \quad h_i^2 + \Psi_i = 1 \\ (ii) \quad Cov(Z_i, F_j) = l_{i1}l_{j1} + l_{i2}l_{j2} + \dots + l_{im}l_{jm}, \quad i, j = 1, 2, \dots, p, \quad i \neq j. \\ (iii) \quad Cov(Z, F) = L_{p \times m} \quad \text{e} \quad \text{portanto,} \quad Cov(Z_i, F_j) = Corr(Z_i, F_j) = l_{ij}, \quad i = 1, 2, \dots, p; \\ j = 1, 2, \dots, m.$$

Assim, pode-se utilizar a matriz  $L_{p \times m}$  na busca do entendimento e interpretação dos fatores  $F_j$ ,  $j = 1, 2, \dots, m$ .

(iv) Em relação à variância total, a proporção explicada pelo fator  $F_j$  é dada por:

$$PVTE_{F_j} = \frac{\sum_{i=1}^p l_{ij}^2}{p} \quad (6)$$

### 2.3.2 TESTE DE ESFERICIDADE DE BARTLETT

Para analisar a matriz de correlação recorreremos ao Teste de Esfericidade de Bartlett, o qual testa a hipótese de que a matriz de correlação das variáveis é uma matriz identidade. Caso não se rejeite  $H_0: P = I$ , significa que  $\rho_{ij} = 0$  ( $i \neq j$ ), então as variáveis não estão correlacionadas.

A estatística do teste é dada por:

$$dr = - \left[ n - 1 - \frac{1}{6}(2p + 5) \right] \log |P| = - \left[ n - \frac{2p + 11}{6} \right] \sum_{j=1}^p \log(\lambda_j) \quad (7)$$

onde  $n$  é o tamanho da amostra,  $p$  número de variáveis e,  $\lambda_j$  ( $j = 1, 2, \dots, p$ ) são os autovalores da matriz de correlação.

Sob a hipótese nula esta estatística de teste tem uma distribuição assintótica  $X^2$  com  $\frac{p(p-1)}{2}$  graus de liberdade. Se o teste de Bartlett toma valores altos de  $X^2$  então rejeita-se  $H_0$  ao nível de significância  $\alpha$ , isto é, as correlações entre as variáveis são significativamente diferentes de zero. Caso não rejeitamos  $H_0$  os autovalores são iguais a 1, seu logaritmo será nulo e a estatística será igual a zero, portanto as variáveis não são correlacionadas, por isso consideramos que a análise fatorial não seja adequada para esses dados.

### 2.3.3 MEDIDAS DE ADEQUAÇÃO

Para medir a adequação da amostra utilizaremos o teste de KMO, proposto por Kaiser, Meyer e Olkin (1958). Este confronta a amplitude das correlações observadas entre as variáveis com a amplitude das correlações parciais, uma vez que esta última aponta o grau de relação entre duas variáveis ignorando a influência das outras. Este índice KMO é dado por:

$$KMO = \frac{\sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij}^2}{\sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij}^2 + \sum_{j \neq i}^p \sum_{i \neq j}^p \rho_{ij(\rho)}^2} \quad (8)$$

que varia entre 0 e 1, quanto menor as correlações parciais mais o índice se aproxima de 1. Segundo Kaiser, Meyer e Olkin (1958),  $KMO \geq 0,75$  é considerado Bom,  $0,5 \leq KMO < 0,75$  aceitável e  $KMO < 0,5$  inaceitável para aplicar a análise fatorial na amostra de dados.

A medida de adequação aqui utilizada para calcular a adequação de cada variável foi a de MSA, similar ao KMO, este índice é dado por:

$$MSA = \frac{\sum_{i \neq j}^p \rho_{ij}^2}{\sum_{i \neq j}^p \rho_{ij}^2 + \sum_{i \neq j}^p \rho_{ij(\rho)}^2}, \quad i = 1, 2, \dots, p \quad (9)$$

Se o resultado de MSA apresentar um valor baixo, podemos detectar as variáveis responsáveis por esse efeito. A análise do MSA é comparável ao KMO, as variáveis encontradas dentro do domínio inaceitável devem ser excluídas para posteriormente aplicar a análise fatorial, ou seja,  $MSA \geq 0,75$  é considerado Bom,  $0,5 \leq MSA < 0,75$  aceitável e  $MSA < 0,5$  inaceitável para proceder a análise fatorial sem a remoção desta variável.

### 2.3.4 ESTIMAÇÃO DO NÚMERO DE FATORES $M$

A escolha do número de fatores é essencial e tem como objetivo a substituição do conjunto de variáveis originais por fatores, sendo natural que o número de fatores seja inferior ao número de variáveis analisadas (CORRAR et al., 2007). Com isso ao invés de se trabalhar com 100% da variabilidade total dos dados, analisa-se somente, uma parcela total dos dados explicadas por fatores.

Segundo Mingoti (2013, pág. 105), o primeiro passo para conduzir a análise fatorial é estimar a matriz de correlação teórica  $P_{p \times p}$ , através da matriz de correlação amostral  $R_{p \times p}$ . Para a estimação de  $m$ , bastará extrair-se os autovalores da matriz  $R_{p \times p}$  e ordená-los de forma decrescente. Observa-se, então, quais autovalores são os mais importantes em termos de grandeza numérica, utilizando os seguintes critérios:

(a) a análise da proporção da variância total relacionada com cada autovalor  $\hat{\lambda}_i$ , dada por  $\hat{\lambda}_i/p$ ,  $i = 1, 2, \dots, p$ . Permanecem aqueles autovalores que representam maiores proporções da variância total e, portanto, o valor de  $m$  será igual ao número de autovalores retidos;

- (b) a comparação do valor numérico de  $\hat{\lambda}_i$  com o valor 1,  $i = 1, 2, \dots, p$ . O valor de  $m$  será igual ao número de autovalores  $\hat{\lambda}_i$  maiores ou iguais a 1. A idéia básica desse critério é manter no sistema novas dimensões que representem pelo menos a informação de variância de uma variável original. Este critério foi proposto por Kaiser (1958);
- (c) usar o gráfico screeplot, representar o valor de  $\hat{\lambda}_i$  ordenados em ordem decrescente, e observar a separação entre os maiores e os menores autovalores. Esse critério é equivalente ao critério (a).

### 2.3.5 MÉTODO DE ESTIMAÇÃO DAS MATRIZES $L_{pxm}$ $\Psi_{pxp}$

Determinado o número de fatores  $m$ , estima-se as matrizes  $L_{pxm}$  e  $\Psi_{pxp}$ . De acordo com o método de componentes principais, para cada autovalor  $\hat{\lambda}_i = 1, 2, \dots, m$ , usado para a estimação do número de fatores  $m$ , encontra-se o auto vetor normalizado correspondente  $\hat{e}_i$ , onde  $\hat{e}_i = (\hat{e}_{i1} \hat{e}_{i2} \hat{e}_{i3} \dots \hat{e}_{ip})'$ . As matrizes  $L_{pxm}$  e  $\Psi_{pxp}$  serão estimadas respectivamente por:

$$\hat{L}_{pxp} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix} \quad (10)$$

$$\hat{\Psi}_{pxp} = \text{diag}(R_{pxp} - \hat{L}_{pxm} \hat{L}'_{m \times p}) \quad (11)$$

Assim, a matriz  $\Psi_{pxp}$  tem a diagonal principal cujos elementos são iguais à diagonal principal da matriz  $(R - \hat{L}\hat{L}')$ . A ideia deste procedimento se encontra na aplicação do teorema de decomposição espectral à matriz  $R_{pxp}$ . Por este teorema, a matriz de correlação amostral pode ser decomposta com uma soma de  $p$  matrizes, cada uma relacionada com um autovalor da matriz  $R_{pxp}$ . Para um valor  $m$  fixo, da seguinte forma:



$$R_{pxp} = \sum_{i=1}^p \hat{\lambda}_i \hat{e}_i \hat{e}'_i = \sum_{i=1}^m \hat{\lambda}_i \hat{e}_i \hat{e}'_i + \sum_{i=m+1}^p \hat{\lambda}_i \hat{e}_i \hat{e}'_i \quad (12)$$

Desse modo, uma aproximação para a matriz  $LL'$  será dada por:

$$\hat{L}\hat{L}' = \sum_{i=1}^m \hat{\lambda}_i \hat{e}_i \hat{e}'_i = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix}', \quad (13)$$

A construção da matriz  $\hat{\Psi}_{pxp}$  é dada pela expressão:

$$\sum_{i=m+1}^p \hat{\lambda}_i \hat{e}_i \hat{e}'_i = R_{pxp} - \hat{L}_{pxm} - \hat{L}'_{m xp} \quad (14)$$

A matriz de correlação amostral original  $R_{pxp}$  é aproximada por:

$$R \approx \hat{L} \hat{L}' + \hat{\Psi} \quad (15)$$

e a matriz residual que serve como um critério para avaliação da qualidade de ajuste do modelo fatorial, é dada por:

$$MRES = R - (\hat{L}\hat{L}' + \hat{\Psi}) \quad (16)$$

### 2.3.6 ROTAÇÃO DOS FATORES

Em alguns casos, a interpretação dos fatores originais  $F_1, F_2, \dots, F_m$  se torna uma tarefa muito difícil devido à aparição de coeficientes  $\hat{l}_{ij}$  de grandeza numérica similar, em vários fatores distintos (MINGOTI, 2013, pág. 119). Caso isso aconteça, é um sinal de que a suposição de ortogonalidade dos fatores foi violada isso implica que, a partição das variáveis originais em fatores se torna difícil de ser compreendida. Esse tipo de situação exige um recurso de transformação ortogonal dos fatores originais. A rotação ortogonal preserva a orientação original entre os fatores, mantendo-os perpendiculares após a rotação (MINGOTI, 2013, pág. 120).

### 2.3.7 CRITÉRIO DE ROTAÇÃO VARIMAX

O método Varimax, proposto por Kaiser (1958), é o método mais utilizado e busca minimizar o número de variáveis que têm altas cargas em um fator, simplificando a interpretação dos fatores (FÁVERO et al., 2009). A busca da matriz  $T_{m \times m}$  tem como base, buscar fatores com grande variabilidade nas cargas fatoriais ou fatores fixos, e um outro grupo de variáveis que tenham baixa correlação ou moderada com o fator. Para cada fator fixo, a solução é obtida através da maximização da variação dos quadrados das cargas fatoriais originais, das colunas da matriz  $\hat{L}_{p \times m}$ . Seja  $\hat{l}_{ij}^*$  o coeficiente da  $i$ -ésima variável no  $j$ -ésimo fator após rotação. Seja  $V$  a quantidade definida por:

$$V = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{l}_{ij}^4 - \frac{1}{p} \left( \sum_{i=1}^p \tilde{l}_{ij}^2 \right)^2 \right] \quad (17)$$

onde  $\tilde{l}_{ij} = (\hat{l}_{ij}^* / \hat{h}_i)$ , sendo  $\hat{h}_i$  a raiz quadrada da comunalidade da variável  $Z_i$ ,  $i = 1, 2, \dots, p$ . A quantidade  $V$  em (17) é proporcional à soma das variâncias das cargas fatoriais escalonados  $\tilde{l}_{ij}$  ao quadrado para cada fator. O escalonamento das cargas fatoriais é feito pelo fato de as variáveis  $Z_i$  não terem necessariamente, comunalidades iguais ou semelhantes.

### 2.3.8 ESTIMAÇÃO DOS ESCORES (MÍNIMOS QUADRADOS PONDERADOS)

Após a identificação e análise dos fatores  $F_j$ ,  $j = 1, 2, \dots, m$ , rotacionados relacionados com as variáveis  $Z_i$ ,  $i = 1, 2, \dots, p$ , será necessário calcular os escores (valor numérico) de cada elemento amostral, de modo a utilizar esses valores em outras análises. Neste trabalho os escores fatoriais serão necessários para a determinação dos pontos da rodovia BR-364 relacionados com cada fator.

Nota-se que o modelo fatorial que foi proposto em (2), é semelhante a um modelo de regressão linear múltipla se considerarmos  $Z$  como um vetor de respostas,  $F$  como vetor de parâmetro do modelo,  $\varepsilon$  o vetor de erros aleatórios com média zero e

matriz de covariâncias  $\Psi$  e  $L_{pxm}$  como a matriz com os níveis das  $m$  variáveis explicativas, ou seja,  $Z = LF + \varepsilon$ . Tratando o modelo fatorial dessa forma pode-se estimar o vetor  $F$  através do método dos mínimos quadrados ponderados (JOHNSON; WICHERN, 2002). Como as matrizes  $L_{pxm}$  e  $\Psi$  não são conhecidas, mas estimadas pelos dados amostrais, tem-se que, para qualquer elemento amostral  $k = 1, 2, \dots, n$ , o respectivo valor numérico no  $\hat{F}_j$  será obtido pela fórmula (18):

$$\hat{F}_{ij} = (\hat{L}' \hat{\Psi}^{-1} \hat{L})^{-1} \hat{L}' \hat{\Psi}^{-1} Z_k = W_{mxp} Z_k \quad (18)$$

onde,  $Z_z = (Z_{1k} Z_{2k} \dots Z_{pk})'$  é o vetor de observações do  $k$ -ésimo elemento amostral padronizado e  $W_{mxp}$  é a matriz de ponderação que gera os coeficientes  $w_{ji}$ ,  $j = 1, 2, \dots, m$ ;  $i = 1, 2, \dots, p$ . Como, as matrizes  $\hat{L}'$  e  $\hat{\Psi}$  foram estimadas pelo método das componentes principais, Johnson e Wichern (2002) recomenda que os escores sejam estimados pelo método dos mínimos quadrados ordinários equivalente ao de mínimos quadrados ponderados.

### 3. METODOLOGIA

Ji-paraná é um município do estado de Rondônia no Brasil. Sua População está estimada em cerca de 131.560 habitantes, sendo o segundo mais populoso do estado e o décimo sexto mais populoso da Região Norte do Brasil, e a 219ª mais populosa do Brasil de acordo com estimativas publicada pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2016).

O município é cortado pela BR-364, rodovia diagonal no sentido sul-norte, ligando Mato Grosso à capital Porto Velho, distante 1.100 km de Cuiabá e 384 km da capital rondoniense, que deu passagem ao desenvolvimento. Por ela, chegam os bens de consumo industrializados do sul do país e saem produtos da terra, através de caminhões de carga.

Ainda que nos últimos anos tenha havido melhoras significativas como, a duplicação da ponte sobre o Rio Machado em 2010 e a inauguração do Viaduto em 2013 para a melhora do trânsito local, Ji-paraná sofre sérios problemas com a BR-364, devido ao intenso movimento de veículos e as precárias condições das estradas, o que

acaba por contribuir com os acidentes rodoviários, principalmente nos locais em que a rodovia corta a cidade.

Os dados aqui utilizados, estão disponíveis no site da Polícia Rodoviária Federal (PRF, 2016), agrupados por ocorrência, e divididos em planilhas, de acordo com cada ano, onde tomou-se para este estudo os anos compreendidos no período entre 01/01/2015 até 31/12/2016. Foram registrados 416 acidentes no período compreendido nos, aproximadamente, 50 quilômetros da BR-364. Utilizando estatística descritiva, será analisado os pontos (KM) com maior número de registros, e posteriormente aplicar a análise fatorial para encontrar os possíveis fatores que estejam relacionadas com esses acidentes.

O banco de dados contém informações sobre o ano do acidente, mês do acidente, dia da semana, KM de ocorrência, causa do acidente (falta de atenção, sono, ingestão de bebidas alcoólicas, não guardar distância de segurança, entre outras), tipo do acidente (capotamento, colisão transversal, colisão traseira, entre outras), classificação do acidente (sem vítimas, com vítimas feridas, com vítimas fatais e ignorado), condições meteorológicas (chuva, céu claro, nublado, entre outras), entre outras.

Apenas uma das variáveis do banco de dados é quantitativa, sendo ela a variável quilômetro (KM), as demais variáveis são qualitativas. Por tanto, para proceder com a análise multivariada de dados dos acidentes de trânsito na BR-364, foi necessário organizar de maneira apropriada dando uma ordem numérica as observações de cada variável. As variáveis qualitativas, que não seguiam uma ordem lógica, foram organizadas de acordo com as frequências relativa, quantificando suas observações desde a menor frequência até a maior, como foi o caso da variável tipo de acidente que não se tem como prever uma ordem lógica, ressaltando que para as variáveis Uso do Solo e Sentido da Via foram mantidas como categóricas para análises posteriores. O resultado dessa organização fica claro na Tabela 1.

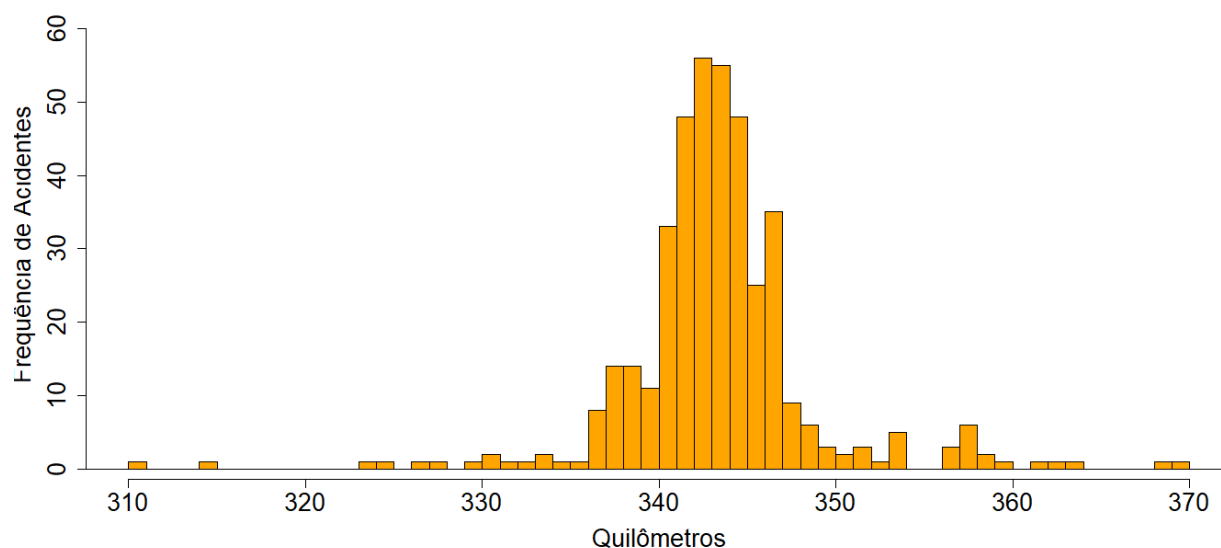
Tabela 1. Tabela de quantificação das variáveis

<b>1. Mês</b>	4. Animais na Pista	3. Com Vítimas Feridas
1. Janeiro	5. Defeito na via	4. Com Vítimas Fatais
2. Fevereiro	6. Ultrapassagem indevida	<b>7. Fase do Dia</b>
3. Março	7. Não guardar distância de segurança	1. Amanhecer
4. Abril	8. Ingestão de álcool	2. Pleno dia
5. Maio	9. Outras	3. Anoitecer
6. Junho	10. Desobediência à sinalização	4. Plena noite
7. Julho	11. Falta de atenção	<b>8. Condições Meteorológicas</b>
8. Agosto	<b>5. Tipo do Acidente</b>	1. Ignorada
9. Setembro	1. Danos Eventuais	2. Chuva
10. Outubro	2. Incêndio	3. Nublado
11. Novembro	3. Colisão com objeto móvel	4. Céu Claro
12. Dezembro	4. Tombamento	5. Sol
<b>2. Dia da Semana</b>	5. Capotamento	<b>9. Tipo de Pista</b>
1. Domingo	6. Atropelamento de animal	1. Simples
2. Segunda-feira	7. Colisão com objeto fixo	2. Dupla
3. Terça-feira	8. Colisão com bicicleta	3. Múltipla
4. Quarta-feira	9. Colisão frontal	<b>10. Traçado da Via</b>
5. Quinta-feira	10. Atropelamento de pessoa	1. Curva
6. Sexta-feira	11. Saída de Pista	2. Cruzamento
7. Sábado	12. Queda de motocicleta / bicicleta / veículo	3. Reta
<b>3. KM</b>	13. Colisão lateral	<b>11. Uso do Solo</b>
Assume valores de: 310 a 369,5	14. Colisão Transversal	Urbano
<b>4. Causa do Acidente</b>	15. Colisão traseira	Rural
1. Dormindo	<b>6. Classificação do Acidente</b>	<b>12. Sentido da Via</b>
2. Velocidade incompatível	1. Ignorado	Crescente
3. Defeito mecânico em veículo	2. Sem Vítimas	Decrescente

Todas essas variáveis serão analisadas de modo a agrupá-las, para uma melhor visualização do perfil dos acidentes ocorridos na BR-364. Para isso foi necessário a utilização de planilha eletrônica para separar e organizar os dados do município em estudo, de cada planilha, uma a uma, pois os dados dispostos contém informações de todos os estados que são cortados por alguma rodovia federal. As análises estatísticas multivariadas, será desenvolvida no software R (R CORE TEAM, 2016).

#### 4. RESULTADOS E DISCUSSÕES

A primeira etapa das análises foi encontrar os quilômetros da rodovia onde se concentram a maior parte dos acidentes. Isso foi possível através do histograma Figura 1 o qual representa as frequências dos acidentes de trânsito ao longo dos quilômetros da BR-364 no período analisado. Ressaltando que pode ter havido quilômetros onde não ocorreu acidente ou, ainda, devida à exclusão de observações com falta de informações, alguns quilômetros deixaram de aparecer nas análises.



**Figura 1. Distribuição dos acidentes entre os quilômetros de 310 ao 370.**

A distribuição de frequência nos fornece uma grande quantidade de informações sobre o conjunto de dados que estamos analisando. Na Figura 1, podemos observar o comportamento e características dos dados como, simetria da distribuição, pois a região

onde há maior concentração de acidentes é o centro da distribuição, a dispersão dos valores encontra-se em torno de um valor central.

Percebe-se que há uma maior incidências de acidentes entre os quilômetros de 335 até 348 aproximadamente, esse intervalo abrange todo o perímetro urbano e corresponde a cerca de 86,2% de todos os acidentes registrados. Mas o destaque na frequência de acidentes está no trecho dos quilômetros 341 ao 343 o qual abrange o perímetro da ponte sobre o Rio Machado que liga o primeiro distrito ao segundo distrito da cidade, este, corresponde a cerca de 33,3% de todos os acidentes registrados. Para melhor explorar as causas dos acidentes nessas regiões, será aplicado a técnica da estatística multivariada de análise fatorial, a qual será desenvolvida logo a diante.

#### **4.1 APLICAÇÃO DA ANÁLISE FATORIAL EXPLORATÓRIA**

Foi necessário aplicar a análise fatorial duas vezes ao conjunto de dados inicial (416 linhas  $\times$  13 variáveis), ressaltando que apenas a segunda análise fatorial foi interpretada. A primeira análise foi realizada com o objetivo de testar a adequacidade da amostra pelo teste de KMO, onde de acordo com Kaiser, Meyer e Olkin (1958),  $KMO \geq 0,75$  é considerado Bom,  $0,5 \leq KMO < 0,75$  aceitável e  $KMO \leq 0,5$  inaceitável, para aplicar a análise fatorial na amostra de dados. E aplicar o teste MSA que tem por objetivo testar a adequacidade das variáveis, eliminando-as de acordo com o critério de aceitação, onde  $0,5 < MSA < 0,75$ .

Para facilitar o desenvolvimento da análise via software, as variáveis foram renomeadas de forma mais simples, conforme a Tabela 2.

Tabela 2. Descrição das variáveis

Variável	Descrição
X1	Mês de ocorrência do acidente
X2	Dia da semana em que ocorreu o acidente
X3	KM de ocorrência do acidente
X4	Causa do acidente
X5	Tipo do acidente
X6	Classificação do acidente
X7	Fase do dia que ocorreu o acidente
X8	Condições Meteorológicas no dia do acidente
X9	Tipo de pista
X10	Traçado da via
X11	Uso do solo
X12	Sentido da via

Após o tratamento dos dados, pode-se avaliar a normalidade univariada e multivariada dos dados dos acidentes na rodovia BR-364 por meio do teste de normalidade Shapiro-Wilk, cujo os resultados não deram normalidade no conjunto de dados. Diante deste resultado, a estimação do modelo fatorial ortogonal se dará pelo método de componentes principais, o qual pode ser aplicado ainda que os dados não apresentem normalidade multivariada. Os resultados da Análise Fatorial e suas matrizes de correlação estimadas dos dados da amostra podem ser encontrados abaixo.

$$R = \begin{bmatrix} 1 & & & & & & & & & \\ 0,02 & 1 & & & & & & & & \\ 0,09 & 0,02 & 1 & & & & & & & \\ 0,08 & 0,05 & -0,01 & 1 & & & & & & \\ -0,01 & 0,20 & 0,07 & 0,27 & 1 & & & & & \\ 0,07 & 0,02 & 0,00 & 0,03 & -0,08 & 1 & & & & \\ 0,01 & 0,04 & -0,06 & -0,15 & -0,13 & 0,10 & 1 & & & \\ 0,11 & 0,07 & 0,03 & 0,10 & 0,08 & 0,04 & -0,30 & 1 & & \\ 0,00 & 0,07 & -0,13 & 0,15 & 0,16 & 0,00 & 0,01 & 0,03 & 1 & \\ 0,13 & 0,02 & 0,02 & -0,04 & 0,02 & 0,00 & 0,01 & 0,01 & -0,05 & 1 \end{bmatrix} \quad (16)$$



Na primeira matriz de correlação dada em 16 foi verificado, se esta é uma matriz identidade através do Teste de esfericidade de Bartlett. Este teste produziu um p-valor menor que 8,984113e-16 a partir da estatística Qui-Quadrado 166,0681 com 45 graus de liberdade. Isso representa que a matriz de correlação não é uma matriz identidade ao nível de sigficância de 5%, ou seja, há uma associação entre as variáveis. Para a mesma matriz o teste KMO para adequação da amostra considerando as dez variáveis, resultou em 0,54, o que é uma medida aceitável para aplicação da Análise Fatorial e a medida MSA, disposta na Tabela 3, para cada variável foram as seguintes:

**Tabela 3. Medida de adequação de cada variável na primeira aplicação da Análise Fatorial.**

Variáveis	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
MAS	0,49	0,50	0,52	0,59	0,57	0,44	0,51	0,53	0,61	0,48

Após a remoção das variáveis X1, X6 e X10 da análise, verificou-se novamente os valores para do Teste de esfericidade de Bartlett. Este por sua vez produziu um p-valor menor que 2,424951e-18 a partir da estatística Qui-Quadrado 133,337 Q-quadrado com 21 graus de liberdade. Logo a nova matriz de correlação dada em 17, não é uma matriz identidade ao nível de significância de 5%, o que confirma novamente, que há uma associação entre as variáveis.

$$R = \begin{bmatrix} 1 & & & & & & & \\ 0,02 & 1 & & & & & & \\ 0,05 & -0,01 & 1 & & & & & \\ 0,20 & -0,07 & 0,27 & 1 & & & & \\ 0,04 & -0,06 & -0,15 & -0,13 & 1 & & & \\ 0,07 & 0,03 & 0,10 & 0,08 & -0,30 & 1 & & \\ 0,07 & -0,13 & 0,15 & 0,16 & 0,01 & 0,03 & 1 & \end{bmatrix} \quad (17)$$

Para a nova matriz o teste KMO agora considerando as oito variáveis, resultou em 0,56 e a medida MSA para cada variável mantiveram os mesmos valores conforme a Tabela 4.

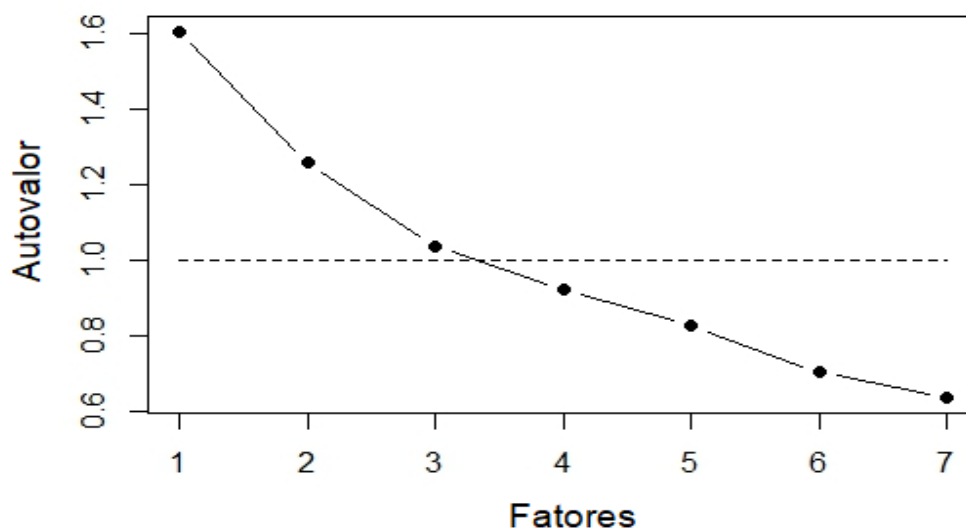
**Tabela 4. Medida de adequação de cada variável na segunda aplicação da Análise Fatorial.**

Variáveis	X2	X3	X4	X5	X7	X8	X9
MAS	0,50	0,51	0,62	0,58	0,53	0,55	0,61

Sendo assim, agora todas as variáveis da análise são aceitáveis segundo ( $0,5 < MSA < 0,75$ ), não havendo a necessidade de eliminar mais nenhuma variável para a aplicação análise fatorial. Feito a aplicação do teste de esfericidade de Barlett e das medidas de adequação da amostra KMO e MSA, e visto que o conjunto de dados é aceitável para o uso da técnica, seguimos com a aplicação da Análise Fatorial.

#### 4.1.2 DETERMINAÇÃO DO NÚMERO DE FATORES

Utilizando o teste Scree, representado na Figura 2 indica que três fatores podem ser mantidos. Levando em conta que, o critério da raiz latente conserva todos os fatores com autovalor maior que um. Estes três fatores representam 56% da variância total dos registros.

**Figura 2. Teste Scree para análise de componentes.**

Dado que o conjunto de dados não possui normalidade, os 3 fatores da análise fatorial foram obtidos pelo método das componentes principais e usando o critério de

rotação varimax, descritos em detalhes abaixo. As cargas fatoriais juntamente com a comunalidade de cada variável são apresentadas na Tabela 5.

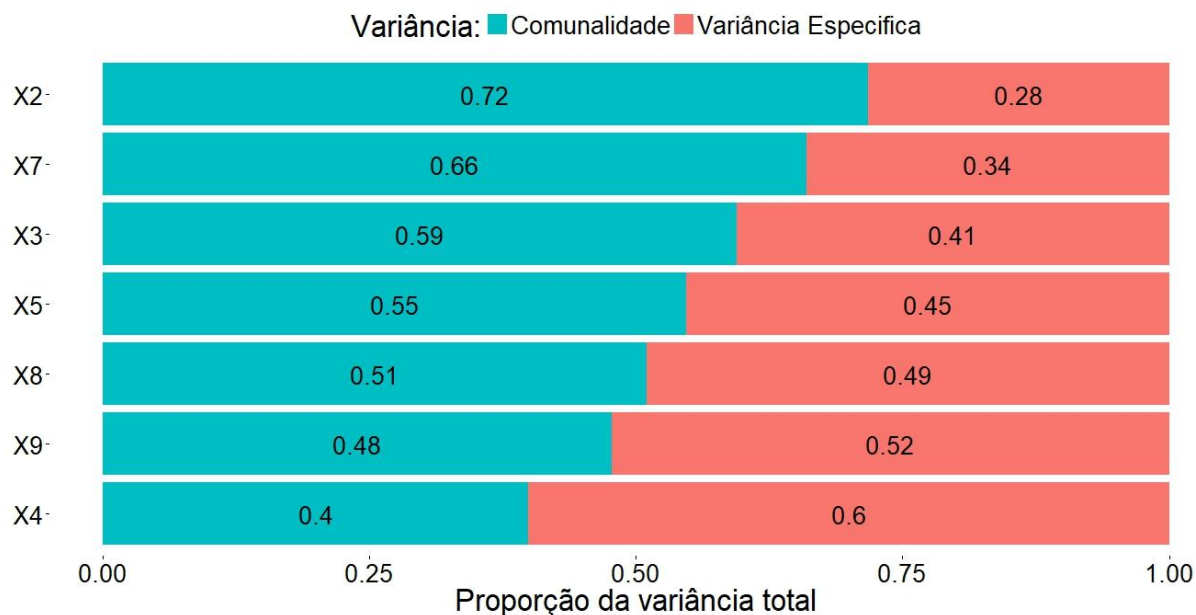
**Tabela 5. Rotação ortogonal da matriz fatorial.**

Antes da Rotação					Depois da Rotação		
Variáveis	Cargas Fatoriais			h2	Varimax		
X2	0,33	0,28	0,73	0,72	-0,15	0,82	-0,16
X3	-0,09	-0,49	0,59	0,59	0,14	0,20	-0,73
X4	0,62	0,07	-0,08	0,40	0,40	0,34	0,36
X5	0,67	0,26	0,18	0,55	0,23	0,62	0,33
X7	-0,49	0,61	0,21	0,66	-0,81	0,06	0,04
X8	0,49	-0,52	-0,08	0,51	0,71	0,06	-0,07
X9	0,41	0,49	-0,26	0,48	0,01	0,23	0,65
Autovalor					1,41	1,27	1,23
Proporção de variância					0,20	0,18	0,18
Variância acumulada					0,20	0,38	0,56
Proporção Explicada					0,36	0,33	0,31
Proporção Acumulada					0,36	0,69	1

Depois de aplicada a rotação fatorial, verificou-se que as comunalidades permaneceram inalteradas após o procedimento conforme a Figura 3, e a quantidade total de variância explicada pelos fatores é a mesma para as soluções rotacionadas e não rotacionadas. Notemos também que, o percentual de variância foi redistribuído de maneira equilibrada pela rotação Varimax, de modo que na solução rotacionada o primeiro, segundo e terceiro fator explicam, respectivamente, 20%, 18% e 18% da variância mantendo os 56% de variância encontrados antes da rotação.

Na Figura 3 está representada a variância comum ou comunalidade e a variância específica de cada variável, vemos que as variáveis X2 (dia da semana) e X7 (fase do dia) são as que mais compartilham suas variâncias com todas as demais variáveis consideradas. Em contra partida as variáveis X4 (causa do acidente) e X9 (tipo de

pista), são que menos compartilham de suas variâncias com as demais, principalmente a variável X4 onde 60% de sua variabilidade é específica e não compartilhada.



**Figura 3. Gráfico da comunalidade e variância especificada.**

#### 4.1.3 INTERPRETAÇÃO DOS FATORES

Para considerar quais cargas são significativas na matriz fatorial, vamos considerar como critério cargas significantes aquelas maiores que 0,5. Após a rotação dos fatores pelo critério de Varimax, o fator 1 permaneceu com 2 variáveis significantes, Fase do Dia (X7) e Condições Meteorológicas no dia do acidente (X8). Para o fator 2 as variáveis significantes foram, Dia da Semana (X2) e Tipo de Acidentes (X5). E para o fator 3, as variáveis significantes foram, Quilômetros de Ocorrência do Acidente (X3) e Tipo de Pista (X9).

A variável Causa do acidente (X4), não apresentou carga fatorial significativa em nenhum dos 3 fatores, mesmo após a rotação dos fatores, além de ter apresentado comunalidade abaixo dos níveis de explicação aceitáveis 0,5. A variável X9 também apresentou comunalidade abaixo dos níveis de explicação, porém, possui carga fatorial significativa no fator 3, como o intuito da análise é a redução de dados a variável X9 foi

mantida na análise. Essa situação talvez pudesse ser corrigida aumentando o número de fatores, porém se aceitado o fator 4 ele não atenderia o critério de autovalor maior que 1. As variáveis incorporadas a cada fator estão apresentadas na Tabela 6.

**Tabela 6. Interpretabilidade dos fatores a partir do critério de rotação Varimax.**

Fator	Variáveis Originais
1	(X7) Fase do Dia (X8) Condições Meteorológicas no dia do acidente
2	(X2) Dia da Semana (X5) Tipo de Acidentes
3	(X3) Quilômetros de Ocorrência do Acidente (X9) Tipo de Pista

O fator 2 apresentou cargas fatoriais positivas em suas variáveis contidas X2 e X5, os fatores 1 e 3, apresentaram cargas fatoriais significantes inversamente correlacionadas. Para o fator 1 as variáveis inversamente correlacionadas foram X7 que apresentou carga negativa e X8 com carga positiva. Já no caso do fator 3 a variável 3 apresentou carga negativa e a variável X9 carga positiva. A explicação das cargas dentro de cada fator ficam interpretadas da seguinte forma, as cargas fatoriais inversamente correlacionadas no fator 1 se explica pelo fato de estas duas variáveis, serem relativamente parecidas, visto que as condições meteorológicas da variável céu claro e sol contidas em X8, sempre irão ocorrer inversamente das fases do dia anoitecer e plena noite contidas na variável fase do dia. Com isso também se percebe que os acidentes registrados ao anoitecer ou em plena noite foram registrados em sua maioria em condições de chuva ou céu nublado. O fator 2 está caracterizado pelas variáveis X2 e X5, estas duas variáveis relacionam o tipo de acidente com o dia da semana, onde de acordo com os resultados obtidos os principais tipos de acidentes como, colisões lateral, traseira e transversal ocorrem na sua maioria nos finais de semana, sexta e sábado. Já os acidentes com menor frequência como incêndio, danos eventuais, atropelamento de animais entre outros tendem a ocorrer no começo da semana. Por último o fator 3 está representado pelas variáveis X3 e X9, que assim

como no fator 1 também são inversamente correlacionadas. Estas caracterizam o fator 3 relacionando o tipo da pista com o quilômetro de ocorrência. Por tanto os registros de acidentes nos quilômetros iniciais do estudo são relacionados com tipo de pista múltipla, o perímetro urbano está compreendido pelo tipo de pista dupla e os registros de acidentes nos quilômetros finais do trecho em estudo, ocorreram em pista simples.

A partir da interpretabilidade dos fatores acima, definimos o fator 1 como "fator diário meteorológico", o fator 2 como "fator decorrente do dia da semana" e o fator 3 como "fator estrutural do quilômetro de ocorrência".

#### 4.1.4 VALIDAÇÃO DO MODELO

A validação do modelo ocorreu por meio da bondade de ajuste. Este método consiste em estimar a matriz residual a partir de diferenças entre as correlações observadas e as correlações reproduzidas. A matriz de correlações observadas é dada pela matriz 17 e a matriz de correlações reproduzidas foi estimada conforme a equação 5 apresentada na subseção 2.3.1. A matriz que contém a variância específica de cada variável  $\Psi$  e utilizada para estimar a matriz reproduzida é apresentada abaixo:

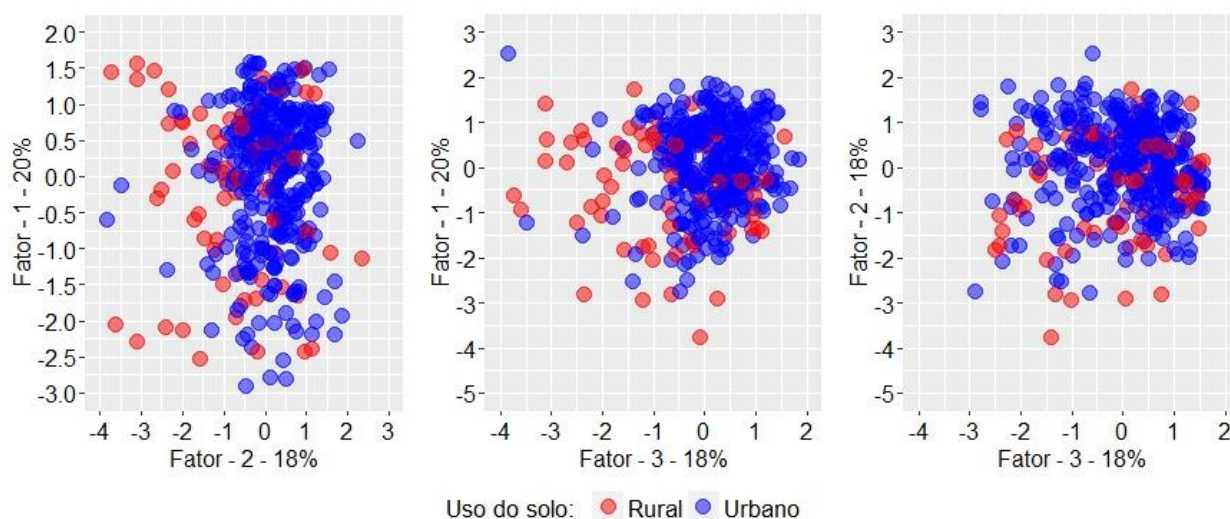
$$\Psi = \begin{bmatrix} 0,28 & & & & & & \\ 0,00 & 0,41 & & & & & \\ 0,00 & 0,00 & 0,60 & & & & \\ 0,00 & 0,00 & 0,00 & 0,45 & & & \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,34 & & \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,49 & \\ 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,00 & 0,52 \end{bmatrix} \quad (18)$$

Portanto a matriz residual é dada por:

$$\text{Matriz residual} = \begin{bmatrix} 0,00 & & & & & & \\ -0,24 & 0,00 & & & & & \\ -0,12 & 0,13 & 0,00 & & & & \\ -0,22 & 0,02 & -0,16 & 0,00 & & & \\ -0,12 & 0,07 & 0,14 & 0,01 & 0,00 & & \\ 0,11 & -0,13 & -0,18 & -0,10 & 0,27 & 0,00 & \\ -0,01 & 0,30 & -0,16 & -0,19 & -0,03 & 0,06 & 0,00 \end{bmatrix} \quad (19)$$

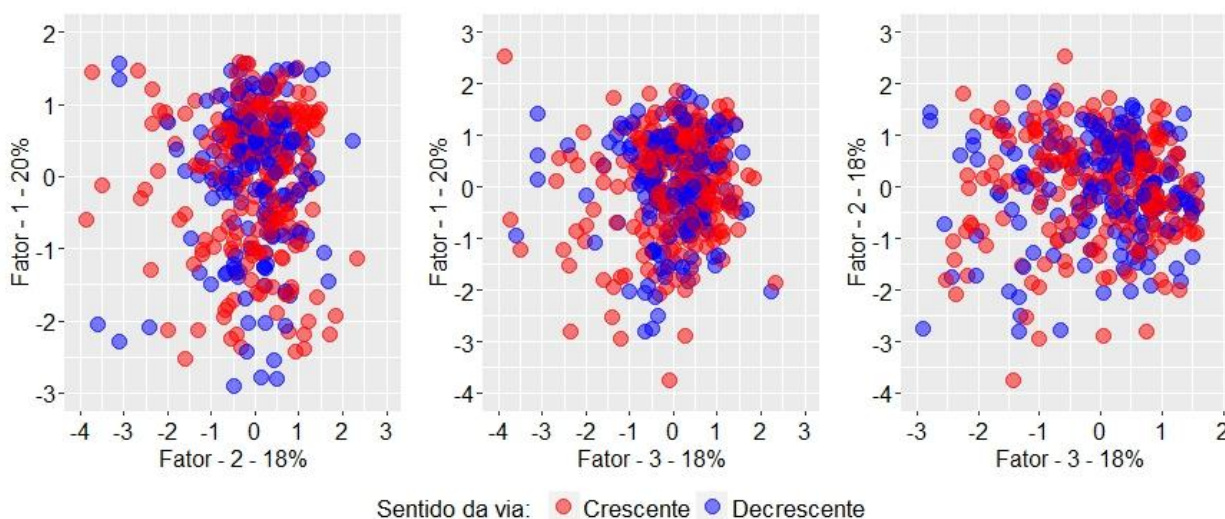
De acordo com a matriz residual dada em 19 temos que aproximadamente 34% dos resíduos são significativos considerando o ponto de corte igual a 0,10. Ressaltamos que essa porcentagem é consideravelmente baixa, isso pode ter ocorrido devido a quantidade de variância explicada pela solução fatorial em 3 fatores ser igual a 56%. Caso fosse considerado a solução fatorial com mais de 3 fatores essa proporção de resíduos significativos aumentaria, porém ao considerar mais de 3 fatores não possivelmente dificultaria uma interpretabilidade prática do problema tratado.

A Figura 4 mostra as projeções dos escores no espaço fatorial relacionando os fatores com a variável Uso do Solo. Fica claro na Figura 4 a grande incidência de acidentes em perímetro urbano e como esses fatores se distinguem em dois grupos.



**Figura 4. Projeção dos escores nos planos fatoriais com relação a variável uso do solo.**

Na Figura 5 as projeções dos scores no espaço fatorial relaciona os fatores com a variável Sentido da Via. Esta por sua vez se mostra de difícil compreensão, pois os fatores não apresentam um padrão ou mesmo grupos, isso implica dizer que o sentido da via, crescente ou decrescente, não tem muita influência com os fatores de ocorrência de acidentes na BR-364.



**Figura 5. Projeção dos escores nos planos fatoriais com relação a variável sentido da via.**

## 5. CONSIDERAÇÕES

A partir da análise realizada, pretendia-se identificar os principais quilômetros de ocorrência dos acidentes, encontrar e caracterizar os fatores que melhor explicassem os registros de acidentes na BR-364 no município de Ji-Paraná no estado de Rondônia. Para tal fim, utilizou-se análise estatística descritiva e Análise Fatorial com o critério Varimax para a rotação dos fatores e validação do modelo fatorial.

Feito todas as verificações necessárias para saber se o conjunto de dados é apropriado, o teste de Esfericidade de Bartlett provou que a matriz de correlação não é uma matriz identidade ao nível de significância de 5%. O teste de KMO assim como a medida MSA resultaram em valores aceitáveis para aplicação da Análise Fatorial, após a remoção das variáveis X1 (Mês de ocorrência), X6 (Classificação do acidente) e X10 (Traçado da via).

A determinação do número de fatores foi através dos critérios da raiz latente e teste Scree, ambos concordantes em seguir a análise com 3 fatores. O critério de rotação Varimax mostrou-se eficaz na interpretação das cargas fatoriais contidas na matriz de cargas fatoriais. Através disso, caracterizaram-se esses três fatores distintos



como, "fator diário meteorológico", "fator decorrente do dia da semana" e "fator estrutural do quilômetro de ocorrência".

Na representação gráfica dos escores, observamos que os pontos não se dispersam de forma aleatória nos planos fatoriais para variável Uso do Solo, mostrando certa tendência para a ocorrência de acidentes em perímetro urbano, diferentemente acontece nos planos para variável Sentido da Via, onde os escores não apresentaram um comportamento tendencioso em relação a algum fator em especial. Isso implica dizer que os acidentes de trânsito da BR-364 na cidade de Ji-paraná acontecem de forma homogênea tanto no sentido crescente quanto no sentido decrescente da rodovia.

Considerando a quantidade de registros disponibilizados pela PRF, podem-se aplicar novos trabalhos utilizando outras técnicas da estatística multivariada como regressão multinomial para encontrar uma distribuição de probabilidade da ocorrência dos acidentes, mesmo em outras cidades cortadas por alguma rodovia federal.

## **6. CONCLUSÕES**

Concluimos que, a técnica de análise fatorial se mostrou eficiente na redução da variabilidade do banco de dados. Aproximadamente 86,2% dos acidentes registrados ocorreram em perímetro urbanos destes, 33,3% ocorreram entre os quilômetros 341 ao 343, respectivamente o trecho que corresponde ao trecho da ponte sobre o Rio Machado.

A Análise Fatorial resultou em três fatores, de um total inicial de 10 variáveis após a rotação dos fatores, "fator diário meteorológico", "fator decorrente do dia da semana", "fator estrutural do quilômetro de ocorrência", representando 56% da variabilidade total.

Com a representação dos escores, encontramos o comportamento dos fatores dentro dos planos. O plano uso do solo resultou em dois grupos bem distintos de ocorrência, com tendência para maior ocorrência em perímetro urbano. Quanto ao plano sentido da via os dados se dispersaram de forma bem homogênea, indicando não haver tendência tanto em sentido crescente quanto no sentido decrescente da via.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

BALBO, F. A. N. et al. **Estudo dos Acidentes na BR-277 Por Meio da Análise Fatorial**. p. 1022–1023, 2010.

BERGAMASCHI, R. B. **SIG aplicado a segurança no trânsito -Estudo de caso no município de Vitória -ES**. 2010.

CORRAR, L. J.; PAULO, E.; FILHO, J. M. D. **ANÁLISE MULTIVARIADA para os cursos de Administração, Ciências Contábeis e Economia**. 1. ed. São Paulo, Atlas/FIPECAFI, 2007.

FÁVERO, L. P. L. et al. **Análise de Dados - Modelagem Multivariada para Tomada de Decisões**. Rio de Janeiro, Campus: Elsevier, 2009.

GOLD, P. A. **Aplicações de engenharia para reduzir acidentes**, Washington, EUA: Inter-American Development Bank, 1999.

HAIR JR, J. F.; ANDERSON, R.E.; TATHAN, R.L.; BLACK, W.C.; BABIN, J.B. **Análise Multivariada de Dados**. Tradução Adonai Schlup Sant' Anna. 6ª ed. Porto Alegre: Bookman, 2009.

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Disponível em: <<https://cidades.ibge.gov.br/brasil/ro/ji-parana/panorama>>. Acesso em: 6 nov. 2016.

IPEA. Acidentes de trânsito nas rodovias federais brasileiras: Caracterização, tendências e custos para a sociedade. **Instituto de Pesquisa Econômica Aplicada (IPEA)**, p. 42, 2015.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3ª ed. New York, Prentice Hall, 2002.

KAISER, H. F. **The varimax criterion for varimax rotation in factor analysis**. **Psychometrika**, v. 23, n. 3, p. 187–204, 1958.

MINERVI, N. A. **Análise Fatorial e de Séries Temporais Aplicadas às Séries dos Acidentes de Trânsito nas Rodovias Federais Brasileiras**. p. 158, 2013.

MINGOTI, S. A. **Análise De Dados Através De Métodos De Estatística Multivariada: Uma Abordagem Aplicada**. Belo Horizonte: Editora da UFMG; 2013.

OMS. **Organização Mundial da Saúde**. Disponível em: <<https://nacoesunidas.org/oms-brasil-e-o-pais-com-maior-numero-de-mortes-de-transito-por-habitante-da-america-do-sul/>>. Acesso em: 6 nov. 2016.

PRF. **Polícia Rodoviária Federal**. Disponível em: <<https://www.prf.gov.br/portal/dados-abertos/acidentes/acidentes>>. Acesso em: 20 set. 2016.

R. **A language and environment for statistical computing, R Development Core Team, R Foundation for Statistical Computing** Vienna, Austria, 2016. Disponível em: <<http://www.r-project.org>>

SEHABER, V. F. **ANÁLISE ESTATÍSTICA MULTIVARIADA DOS ACIDENTES DE TRÂNSITO DA BR-376 NO PERÍODO ENTRE OS ANOS DE 2009 E 2012**. 2013.

SPEARMAN, C. “ **General Intelligence** ,” Objectively Determined and Measured. v. 15, n. 2, p. 201–292, 1904.